CHAPTER 7

Initial Calibration

7-1.   Introduction.

The **initial calibration** is evaluated to ensure that the instrument was capable of producing acceptable quantitative data *prior* to the analysis of environmental samples. The concentration range and number of standards for the initial calibration will be dependent upon the instrument, method, and objectives of the project. The variation in instrumental response with concentration may define a straight line or some curve. Instrumental response may be expressed either as peak area (e.g., determined from a sum of detector signals) or as peak height (e.g., the maximum detector signal above background noise).

7-2.   Acceptance Criteria.

The calibration criteria presented below are based upon the guidance presented in USACE Shell document and SW-846 (Method 8000B). However, the criteria are conservative and should be applicable to most instrumental methods. Calibration may be performed using linear or nonlinear fits. However, linear calibrations should be used in preference to nonlinear calibrations

7-2.1. Frequency.

   a. An initial calibration must be performed prior to the analysis of samples and when a continuing calibration verification is unacceptable.

   b. For inorganic analyses, the initial calibration is typically performed at the beginning of each analytical shift in which analyses are performed; that is, each time the instrument is set up to perform analyses (e.g., turned on and "warmed up"). When analyses are performed continually, the initial calibration is typically performed on a daily basis (i.e., every 24 hours).

7-2.2. Number of Calibration Standards.

   a. The number of calibration standards (or points) is highly method dependent. The number of calibration standards will be proportional to the variability of instrumental response. For example, the higher the variability of response, the greater the number of calibration standards that will be required. As a general rule, at least three calibration standards should be used for linear calibrations. Additional calibration standards should be used if the linear calibration range is greater than one or two orders of magnitude or when nonlinear calibrations are performed.

   b. When calibration curves are fitted with polynomials using regression analysis, the number of initial calibration standards must be sufficient for at least one statistical degree of freedom. The degrees of freedom for a regression curve, $df = n - 1 - k$, where $k > 0$. The variable k is the order of the polynomial and n is the number of initial calibration standards. For example, if a regression line is being used ($k = 1$), then the number of calibration standards (n) must be

greater than or equal to three. However, beyond the minimum number of standards required to perform the actual mathematical fit (e.g., three standards for linear regression lines), the number of initial calibration standards required is somewhat arbitrary: Any number standards may be used to generate a calibration curve that meets the specified tolerance for uncertainty. However, when regression analysis is not being performed, only a single calibration standard may be required. For example, if response has been demonstrated to be linear through the origin (e.g., as in ICP analyses), the initial calibration may be performed using only a single high-level standard.

c. It is recommended that the number of standards be equal to or greater than that specified in the most applicable promulgated USEPA method. For example, if SW-846 methods are specified in the Sampling and Analysis Plan (SAP), then the guidance presented in the individual SW-846 method should be followed *to the extent that is appropriate for the objectives of the project*. For the purposes of illustration, the guidance in Update III of SW-846 for linear calibrations is summarized below. However, additional criteria are also specified.

(1) At least five calibration standards are used for all single-component target analytes and surrogates. (Single-standard calibrations are not typically acceptable for surrogates).

(2) A blank and a minimum of three calibration standards are used for SW-846 trace metal analyses (e.g., GF-AA, FL-AA, and ICP analyses). However, a blank and a minimum of five calibration standards are used for mercury analyses.

(3) A blank and one calibration standard may be used for ICP analyses. However, though not specified in Method 6010B, *the initial calibration line should be initially verified using a low-level CCV in order to establish the method quantitation limit.*

(4) Calibration for multi component analytes such as Aroclors are performed using a single standard. Instrumental response is related to concentration using the peak areas or heights of several peaks (e.g., at least three **characteristic peaks** for the Aroclors). *However, contrary to the guidance presented in the method, a minimum of three calibration standards is recommended for chromatographic analyses for multi-component analytes.*

7-2.3. Linear Calibration.

Linear calibrations must be performed using regression analyses (with the possible exception of chromatographic analyses). Instrumental response, as measured by the slope of the calibration line, must be high relative to analytical uncertainty (e.g., calibration lines with very "shallow" slopes would not acceptable).

7-2.3.1. Organic Methods.

The coefficient of determination (i.e., the square of the correlation coefficient r) must be equal to or greater than *0.980* for each target analyte. Mean response factors may be used to perform linear calibrations through the origin for chromatographic analyses. The **percent relative standard deviation (%RSD)** for the mean response factors must be equal to or less than *15%* for each target analyte. However, if only three calibration standards are used (e.g., for multi compo-

nent analytes such as PCBs), the coefficient of determination should be *0.990* or greater or the %RSD should be *10%* or less. In addition, any method-specified minimum response factors must be met.

> Note: A number of analytical methods (e.g., SW-846 methods) specify a maximum %RSD of 20% for the initial calibration for chromatographic methods with 2-D detectors. An acceptance limit of 20% is not recommended. Initial calibration lines with %RSDs of 20% often exhibit poor linear fits (e.g., a calibration line may not adequately fit the data at the upper end of the calibration range or the regression coefficient may be less than 0.99).

7-2.3.2. Inorganic Methods.

The coefficient of determination (square of the correlation coefficient r) must be at least *0.990* for each target analyte. Note that when a single standard is used to perform the initial calibration for ICP analyses, a correlation coefficient cannot be calculated. Under these circumstances, the initial calibration must be evaluated using CCVs at multiple concentrations.

7.2.4. Nonlinear Calibration.

a. Nonlinear calibrations are appropriate when linear calibrations cannot be performed over a sufficiently wide working range (e.g., when detector response is inherently nonlinear over a calibration range that spans two orders of magnitude or less). Nonlinear calibrations are inappropriate to compensate for detector saturation at higher concentrations or to avoid proper instrument maintenance. A large reduction in instrumental response (e.g., "curve flattening" characteristic of detector saturation) must not occur near the upper portion of the curve corresponding to high concentrations. Instrumental response must be high relative to analytical uncertainty and must be similar in magnitude to that for linear calibration.

b. All nonlinear calibrations must be performed using regression analysis. Nonlinear calibration curves must be generated using polynomial fits of no higher than third order (e.g., $y = ax^3 + bx^2 + cx + d$) and must possess at least three statistical degrees of freedom. The coefficient of determination must be equal to or greater than *0.99*.

7-2.5. Intercept of Calibration Curve.

A calibration curve (whether linear or nonlinear) must *not* be forced through the origin unless it is demonstrated (e.g., during method development) that the intercept (i.e., $y[x = 0]$) is not statistically different from zero (e.g., by performing a t-test for the y-intercept or comparing it to the MDL.) Arbitrarily forcing a calibration curve through the origin may adversely impact low-level quantitative results. Similarly, when calibration curves are generated using regression analysis, the curves must not be artificially weighted toward the origin by including the point (0,0) one or more times.

7-3.   Evaluation.

Review the standard preparation and run sequence log sheets to verify that the initial calibration was performed at the appropriate frequency using the appropriate number of standards. Review the calibration summary results (e.g., Chapter 4.2.10) to ensure that acceptable initial calibrations, CCVs, and ICVs were performed.

7-3.1. "Goodness of Fit."

a. Check the linearity or "goodness-of-fit" for the calibration curve for each target analyte and surrogate by performing a qualitative visual examination of each calibration plot. Any unusual problem observed during the visual examination of the calibration plots must be noted (e.g., shallow slopes, high curvature at the upper of the calibration curve indicative of detector "saturation," large negative or positive intercepts, and large scatter).

b. When regression analysis is performed, it is especially critical to evaluate the fit for the calculated curve near the origin. In particular, a high correlation coefficient can be obtained (e.g., r > 0.995) when the calibration points significantly deviate from the curve.

c. Determine whether each of the reported regression coefficients or %RSDs are acceptable. Using the calibration summary forms, recalculate the regression coefficient or %RSD (depending on how the initial calibrations were performed) for at least one target analyte, and compare the recalculated value to the reported value. If calibration was performed using the **internal standard technique,** recalculate the relative response factors, mean response factor and %RSD for one target analyte using the appropriate internal standard. The reported and calculated values must agree to within *two* significant figures.

7-3.2. Representativeness of Initial Calibration Curve.

a. Verify that the initial calibration is representative of the analysis. In particular, review the calibration summary forms to determine if any calibration points were erroneously "censored;" that is, whether any calibration standards were *inappropriately* omitted from the "goodness-of-fit" calculations (e.g., for the calculation of the %RSD or correlation coefficient). Since instrumental response will be linear over a sufficiently narrow concentration range, the standards at the extreme upper and lower ends of the calibration range (i.e., the highest and lowest calibration standards) may be dropped to obtain a better linear fit. *However, it is usually inappropriate to drop calibration standards in the middle of the concentration range!*

b. When any calibration standard between the highest and lowest initial calibration standards is omitted, verify that a legitimate reason is documented (e.g., operator blunder, instrument malfunction, and the inadvertent use of an expired calibration standard). Censoring calibration results (with the exception of the points corresponding to the extreme ends of the concentration range) for no other reason than to obtain a "better" curve fit is an *inappropriate* laboratory practice. An "errant" calibration result may actually be a valid result that reflects higher-than-expected random analytical error. Under these circumstances, the omission of the calibration result will under estimate the actual uncertainty associated with the environmental samples.

7-4. Qualification.

7-4.1. Frequency and Number of Standards.

If the instrument was not calibrated at the appropriate frequency with the appropriate number of standards, qualify the associated sample results using professional judgement. For example, if only four points were used to perform the initial calibration for the mercury analyses and a high correlation coefficient was obtained, the associated mercury results would not typically be rejected. The width of the calibration range and the goodness of fit must be taken into account. As a "rule of thumb," at least three calibration standards are required when the quantitation range is two orders of magnitude in width; additional standards are frequently required for larger calibration ranges.

7-4.2. Representativeness.

When mid-level standards are inappropriately dropped, if possible, recalculate the calibration curve and the "goodness of fit" (e.g., r^2 or %RSD). If the "goodness of fit" is unacceptable, then qualify the associated sample results as discussed in Chapter 7.4.3. If it is not possible or practical to recalcuate the curve and the "goodness of fit," then, at a minimum, qualify all associated detections with the J flag and all nondetections with the UN flag. The X flag may be appropriate if the data are being used to support critical decisions.

7-4.3. "Goodness of Fit."

a. Results must be qualified on the basis of quantitative acceptance limits for the calibration fits (e.g., the regression coefficients) and the *visual examination of the initial calibration plots*. In particular, it may be necessary to qualify results when an *acceptable* coefficient of determination or %RSD is obtained for the calibration. For example, if a nonlinear curve possesses an acceptable coefficient of determination, it would be appropriate to qualify high concentration samples if severe curve "flattening" from detector saturation were observed. Detections near the intercept may be unreliable when a large nonzero y-intercept is obtained.

b. A conservative approach is recommended when method-specified minimum response factors are not met. It is recommended that detections and nondetections be qualified with the X flag or the R flag. (Minimum relative response factors for GC/MS analyses are typically 0.05 or greater.) However, use professional judgement, and, at a minimum, qualify nondetections with the UN flag and detections with the J flag.

c. Table 7-1 lists (to three significant figures) the maximum linear correlation coefficient that is not significantly greater than zero as a function of the number of calibration points (*n*) and confidence level (CL). For example, when the initial calibration is performed using five calibration points (e.g., for organic analyses), a correlation coefficient of 0.93 or less does not demonstrate a positive linear correlation exists between instrument response and concentration at the 99% confidence level. Therefore (using the 99% confidence level), if the correlation coefficient is less than or equal to 0.93, then detections and nondetections are qualified with the X or R flag. Similarly, when the initial calibration is performed using four calibration points (e.g., three cali-

bration standards and a blank for the metal analyses), a positive linear correlation is not demonstrated at the 99% confidence level when the correlation coefficient is less than or equal to 0.98.

Table 7-1
Maximum *r* (Linear Correlation Coefficient) Not Significantly Greater Than Zero Versus Number of Calibration Points for Initial Calibration [1]

n	90% CL	95% CL	97.5% CL	99% CL
3	0.951	0.987	0.996	0.999
4	0.800	0.900	0.950	0.980
5	0.687	0.805	0.878	0.934
6	0.608	0.729	0.811	0.882
7	0.550	0.669	0.754	0.832

Notes: 1. CL denotes the confidence level and *n* denotes the number of points used to generate the initial calibration line. The above table was generated using a one-tail t-test for the null hypothesis that the correlation coefficient is not greater than zero

d. In the absence of project-specific requirements to do otherwise, qualify the results as discussed below. It is assumed that at least five points (i.e., calibration standards) are used to perform the initial calibration. When the initial calibration is performed using less than five points, the acceptance limits for the %RSD and the correlation coefficient should be more stringent.

e. When the initial calibration is performed using five or more calibration points, the acceptance criteria for the initial calibration are not met, and the correlation coefficient is *greater than 0.93 or* the %RSD is *less than or equal to 30%* (when mean response factors are used to perform the initial calibration), then, at a minimum, qualify nondetections with the UN flag and detections with the J flag (as illustrated in Table 7-2). If the correlation coefficient is less than or equal to 0.93 or the %RSD is greater than 30%, then qualify the results (i.e., the detections and nondetections associated with the noncompliant target analyte) with the R flag. However, professional judgement should be applied. For example, if the %RSD is grossly unacceptable, but a review of the response factors for the initial calibration indicates that the %RSD is unacceptable because of high variability at the high-end of the calibration line, it may be more appropriate to qualify nondetections with the UN flag rather than the R flag.

f. Qualification of results for unacceptable calibration fits may be avoided when it is possible to recalculate the calibration curve from the information enclosed in the data package. For example, if the coefficient of determination or %RSD is unacceptable for a linear calibration (e.g., because of curvature at the extreme low or high end of the calibration range), then dropping the high or low end calibration standard may give an acceptable calibration line (i.e., an acceptable fit). The new calibration line may then be used to recalculate all the reported sample results. However, at least three calibration standards are required to calculate a calibration line

using regression analyses. Furthermore, since this strategy narrows the quantitation range, results that fall outside of the new calibration range would need to be qualified (as described below).

> Note: The degree to which recalculations will be required for the initial calibrations to avoid data qualification will be highly situation-dependent. For example, it will be function of the completeness of the data package, the level of effort negotiated for the data review, and the degree to which QC requirements and corrective actions were specified in the contract for laboratory analytical services.

Table 7-2
Data Qualification for "Goodness of Fit" for Organic Analyses with Five-Point Linear Calibrations [1]

"Goodness of Fit"	Remarks	Sample (y)	Flag
$r > 0.99$ %RSD $\leq 15\%$	Acceptable Calibration	$MRL < MQL < y$	None
		$MRL < y < MQL$	J
		$y < MRL$	U
$0.93 < r < 0.99$ $15\% < \%RSD \leq 30\%$	Marginal Failure	$y > MRL$	J
		$y < MRL$	UN
$r \leq 0.93$ %RSD $> 30\%$	Gross Failure	$y > MRL$	R
		$y < MRL$	R

Notes: 1. %RSD, r, MRL, MQL, and y denote the percent relative standard deviation of the response factors, the linear correlation coefficient, method quantitation limit, method reporting limit, and the concentration of analyte in the associated field sample, respectively. (The MRL is assumed to be greater than the MDL but less than the MQL.)

7-4.4. Calibration Range.

a. All detections that fall outside of the calibration range of the instrument must be qualified. For calibrations performed with multiple standards, the lowest and highest calibration standards determine the lower and upper limits of the calibration range, respectively.

b. If the initial calibration is performed using a single (high) calibration standard for ICP analyses, then the lower limit of the calibration range must be established by a low-level CCV standard that is analyzed immediately after the initial calibration. If the initial calibration was not verified via the analysis of a low-level CCV standard or if a low-level CCV was analyzed but possesses a recovery that does not fall within 85% to 115%, then multiply the MDL by a factor of ten (10) to establish the lower calibration limit (and the MQL). However, if the low-level CCV is at least ten times greater than the MDL and the recovery does not fall within 85% to

115%, then, at a minimum, qualify all detections between the low-level and mid-level CCVs as estimated and qualify any associated nondetections with the UN flag.

c. As stated in Chapter 6, detections that are less than the lower limit of the calibration range (e.g., the low calibration standard) must be qualified with the J flag. In addition, all detections that *marginally* exceed the high calibration standard must be qualified with the J flag; detections that *grossly* exceed the upper calibration range must be qualified with the X flag (or the R flag). The criteria for *marginal* versus *gross* failures must be determined using professional judgement. The determinative technique as well as the range, fit, and shape of the calibration curve must be taken into consideration. In general, if a sample result exceeds the upper calibration standard within the uncertainty tolerance for the CCV, it is recommended that the result be qualified as a marginal failure. For example, if the high calibration standard is 100 ppb and the CCV must be within 15% of its expected value, then detections greater than 100 ppb but less than 115 ppb should be qualified with a J flag.